

COMPETITION BETWEEN ‘WHO’ AND ‘WHICH’
 IN SLAVIC RELATIVE CLAUSES

(1) Mikhail Bulgakov. The Master and Margarita (RNC¹)

| | | | | | |
|----|---|----------------------|----------------------|------------------------|------------|
| a. | <i>te,</i> | <i>kto</i> | <i>videl ego</i> | <i>vpervye...</i> | RUSSIAN |
| | <i>tyja,</i> | <i>xto</i> | <i>bačyŭ jago</i> | <i>ŭpersynju</i> | BELARUSIAN |
| | <i>Ty,</i> | <i>kdo</i> | <i>ho viděli</i> | <i>poprvé</i> | CZECH |
| | those | who | saw him | first | |
| b. | <i>ti,</i> | <i>jaki</i> | <i>bačili jogo</i> | <i>vperše...</i> | UKRAINIAN |
| | <i>ci,</i> | <i>którzy</i> | <i>widzieli go</i> | <i>po raz pierwszy</i> | POLISH |
| | <i>oni</i> | <i>koji</i> | <i>su ga vidjeli</i> | <i>prvi put</i> | CROATIAN |
| | those | which | saw him | first | |
| | ‘Those who saw him for the first time...’ | | | | |

1. Slavic ‘who’ vs. ‘which’

- Two types of pronouns among the relativizers of most Slavic languages:
 - ‘**who**’: a relativizer derived from an interrogative pronoun with the meaning ‘who’, which refers to people, does not decline for number and gender, and cannot be used attributively;
 - ‘**which**’: a frequent relativizer, non-sensitive to animacy, which can under some conditions be used attributively and declines for number and gender.
- NB: The terms are conventional:
 - English *who* would fall into the ‘who’ group;
 - English *which* does not quite suit the set of features of the ‘which’ group.
- The distinction of these two groups is relatively consistent across Slavic languages, see Tables 1 and 2 (based on Křížková 1970; Gołąb, Friedman 1972 and grammatical descriptions of individual languages);
- Macedonian *koj* and *kojšto*², Bulgarian *kojto*:
 - more like ‘which’, but the stem is used as the interrogative ‘who’ (as well as ‘what’), see Table 2;
 - do not have a counterpart belonging to the ‘who’ group;
 - are not considered below in any detail.

Table 1. Slavic ‘who’

| | | only human | stem = interrogative ‘who’ = ‘who’ | stem = ‘who’ + smth. | no number or gender distinctions | cannot be used attributively |
|--------------|----------------|--------------|------------------------------------|----------------------|----------------------------------|------------------------------|
| East Slavic | Russian | <i>kto</i> | + | + | + | + |
| | Ukrainian | <i>xto</i> | + | + | + | + |
| | Belarusian | <i>xto</i> | + | + | + | + |
| West Slavic | Polish | <i>kto</i> | + | + | + | + |
| | Czech | <i>kdo</i> | + | + | + | + |
| | Slovak | <i>kto</i> | + | + | + | + |
| | Upper Sorbian | <i>štóž</i> | +/- ³ | | + | ? ⁴ |
| | Lower Sorbian | <i>chtož</i> | +/- | | + | ? |
| South Slavic | Slovene | <i>kdor</i> | + | + | + | + |
| | Serbo-Croatian | <i>(t)ko</i> | + | + | + | + |

¹ Russian National Corpus, ruscorpora.ru.

² The latter is very infrequent in the corpus.

³ According to (Bartels, Spiess 2012), relative ‘who’ can refer to inanimates in older literary Lower Sorbian and non-standard Upper Sorbian, unlike the present-day standard languages.

⁴ According to (Bartels, Spiess 2012), the interrogative *chto* has number forms and it could be true for relative pronouns.

Table 2. Slavic ‘which’

| | | | non-sensitive to animacy | stem ≠ interrogative ‘who’ | has number and gender distinctions | can be used attributively |
|-----------------|----------------|----------------|-----------------------------|-------------------------------|--|------------------------------|
| East Slavic | Russian | <i>kotoryj</i> | + | + | + | + |
| | Ukrainian | <i>kotrij</i> | + | + | + | + |
| | | <i>jakij</i> | + | + | + | + |
| | Belarusian | <i>katory</i> | + | + | + | + |
| | | <i>jaki</i> | + | + | + | + |
| West Slavic | Polish | <i>który</i> | + | + | + | + |
| | Czech | <i>který</i> | + | + | + | + |
| | Slovak | <i>ktorý</i> | + | + | + | + |
| | Upper Sorbian | <i>kotryž</i> | + | + | + | + |
| | Lower Sorbian | <i>kótaryž</i> | + | + | + | + |
| South Slavic | Slovene | <i>kateri</i> | + | + | + | + |
| | Serbo-Croatian | <i>koji</i> | + | +/- ⁵ | + | + |
| | | <i>koj</i> | + | - | + | + |
| | Macedonian | <i>kojšto</i> | + | - | + | + |
| Bulgarian | <i>kojto</i> | + | - | + | + | |

- The features in the columns need not pattern together, e.g.
 - English *which*
 - ≠ interrogative ‘who’;
 - does not have distinct number forms;
 - French *qui*
 - = interrogative ‘who’;
 - insensitive to animacy as a subject relativizer;
- However, as in most Slavic languages these features go together, as for now, I will consider all these features as parts of definitions, allowing for some minor deviations, marked in the tables.
- Thus defined ‘who’ and ‘which’ in other languages:
 - ‘who’: German *wer*; Mordvin-Moksha *kijə*...
 - ‘which’: Finnish *joka*; Mordvin-Moksha *kona*...

2. The data

2.1. The basics

- The most systematic overview so far (to the best of my knowledge) is given by Křížková (1970);
- Most of her findings can be summarized as follows:

Table 3. The use of ‘who’ and ‘which’ in Slavic languages, according to the data in Křížková (1970)

| | questions | correlatives | ‘someone’, ‘nobody’, ‘each’ | ‘that’ | ‘all’, ‘those’ | nouns in the head |
|-----------------------------|-----------|--------------|--------------------------------|-----------|----------------|----------------------|
| Russian | who | who/which | who | who/which | who/which | who/’which |
| Ukrainian | who | who | who | who/which | who/which | who/’which |
| Belarusian | who | who | who | who/which | who/which | who/’which |
| Polish | who | who | who | who/which | which | which |
| Czech | who | who | who | who/which | who/which | which |
| Slovak | who | who | who/which | who/which | which | which |
| Slovene | who | who | which | who/which | which | which |
| Upper Sorbian ⁶ | who | who/which | which | who/which | which | which |
| Lower Sorbian | who | who | (no data) | who | which | which |
| Serbo-Croatian ⁷ | who/which | who/which | who/which | who/which | who/which | which |

⁵ Serbo-Croatian *koji* can be used as an interrogative pronoun.

⁶ ‘who’ is attested more widely in older literary Lower Sorbian and non-standard Upper Sorbian (Bartels, Spiess 2012: 227).

- (2) RUSSIAN, questions
Kto iščet?
 who looks_for
 ‘Who searches?’
- (3) RUSSIAN, correlative
Kto iščet, tot najdet...
 who looks_for that will_find
 ‘The one who searches will find’.
- (4) RUSSIAN, ‘each’
Každyj, kto iščet, najdet...
 each who looks_for will_find
 ‘Everyone who searches will find’.
- (5) RUSSIAN, ‘that’
Tot, kto iščet, najdet...
 that who looks_for will_find
 ‘The one who searches will find’.
- (6) RUSSIAN, ‘these’
Te, kto iščet, najdut...
 these who looks_for will_find
 ‘Those who search will find’.
- (7) RUSSIAN, noun in the head
Ljudi, kotorye iščut, najdut...
 people who look_for will_find
 ‘People who search will find’.
- Not all the data is given in Table 3:
 - The ordering of the variants within a language is sometimes mentioned, but not quite systematically, so I leave it out in all the cases;
 - There’s some more sketchy data on the distribution with
 - ‘first’, ‘last’, etc.;
 - ‘that.F’;
 - personal pronouns;
 - cleft constructions.
 - Important **correction**:
 - Upper Sorbian does allow ‘who’ to be used with indefinites (Faßke 1981), cf. also:
- (8) UPPER SORBIAN, Nikolai Ostrovsky. How the steel was tempered (Parasol)
Ale bydenje běše prózdne, a nihtó tu njeběše, kohož by so woprasať
 but flat was empty and nobody there wasn’t whom would ask
 ‘But the flat was empty, and there was nobody to ask’.
- In Russian, *kotoryj* in correlatives
 - only occurs in colloquial language;
 - is probably preferred in the plural, which is impossible for the groups ‘someone’, ‘nobody’, ‘each’; ‘that’. Number is a different parameter (see below), therefore it is possible to come up with the following **interim summary**:
- (9) questions > correlatives > ‘someone’, ‘nobody’, ‘each’; ‘that’ > ‘all’, ‘those’ > heads with nouns

2.2. More data on this part

- The relevant part of the data in Table 3:

⁷ Browne’s (1986: 34) judgements on Serbo-Croatian are in many respects different.

Table 4. Grammaticality of ‘who’ in different contexts, according to the data in Křížková (1970)

| | ‘that who’ | ‘those who’ | ‘all who’ |
|----------------|------------|-------------|-----------|
| Russian | + | + | + |
| Czech | + | + | + |
| Slovene | + | – | – |
| Ukrainian | + | + | + |
| Belarusian | + | + | + |
| Polish | + | – | – |
| Slovak | + | – | – |
| Serbo-Croatian | + | + | + |
| Lower Sorbian | + | – | – |
| Upper Sorbian | + | – | – |

stands for “there’s a claimed difference in acceptability”

- New data based on a corpus study:
 - if a pronoun is allowed in both contexts;
 - Russian of the 18th century.
- I do not consider modern Russian data, because
 - *kotoryj* ‘which’ in modern Standard Russian is used almost exclusively with nouns in the head (including elliptical contexts);
 - *kotoryj* ‘which’ in Non-Standard Russian does occur in these contexts, see, e. g., Spencer (1993), but, as it is very colloquial, some different tendency might be at play.
 - (To put it in other words, Russian colloquial data does not quite seem to fit in. I don’t know why. May be, because of some sociolinguistic factors.)

Table 5. Quantitative data on the use of ‘who’ in different contexts

| | ‘that who’ | ‘those who’ | ‘all who’ | Corpora |
|------------------------|------------------|-------------|-----------|------------------------------|
| Russian of the 18th c. | 0,7 | 0,04 | 0,3 | RNC, main |
| Czech | 0,9 | 0,5 | 0,8 | Intercorp |
| Slovene ⁸ | 1,0 | 0,0 | 0,5 | FidaPLUS |
| Ukrainian | 0,9 | 0,8 | 1,0 | RNC, parallel |
| Belarusian | 1,0 | 0,8 | 0,9 | Belarusian N-corpus, fiction |
| Polish | 0,6 ⁹ | 0,0 | 0,0 | |
| Slovak | 0,6 | 0,0 | 0,0 | Intercorp |
| Serbo-Croatian | 0,4 | 0,0 | 0,0 | |
| Lower Sorbian | 0,1 | 0,0 | 0,0 | DOTKO |
| Upper Sorbian | 0,0 | 0,0 | 0,0 | HOTKO |

stands for “the difference between the two groups is statistically significant” with $p < .05$.

The numbers correspond to the ratio of ‘who’ in these contexts.

For Russian of the 18th c., Czech, and Slovene there’s also statistically significant difference between ‘that’ and ‘all’.

- Some examples of data sets:

⁸ Both strategies are relatively marginal in Slovene, while the preferred option is the indeclinable relativizer.

⁹ The ratios in italics can in fact be higher. The relevant data was not checked manually, because it could not reduce statistical significance.

- Czech:

Table 6. Frequency of ‘who’ and ‘which’ in Czech (Intercorp)¹⁰

| | ‘who’ | ‘which’ | ratio of ‘who’ |
|---------|-------|---------|----------------|
| ‘that’ | 91 | 9 | 0,9 |
| ‘all’ | 78 | 22 | 0,8 |
| ‘those’ | 54 | 46 | 0,5 |

- Russian of the 18th century:

Table 7. Frequency of ‘who’ and ‘which’ in Russian of the 18th century (RNC)¹¹

| | ‘who’ | ‘which’ | ratio of ‘who’ |
|---------|-------|---------|----------------|
| ‘that’ | 206 | 83 | 0,71 |
| ‘all’ | 3 | 7 | 0,30 |
| ‘those’ | 6 | 161 | 0,04 |

The second interim summary:

- (10) questions > correlatives > ‘someone’, ‘nobody’, ‘each’; ‘that’ > **‘all’** > **‘those’** > heads with nouns

3. Discussion

- ? Is there a more exact generalization?
- ? Which properties of the heads and relative clauses really matter for the choice?
- ? Which properties of relativizers trigger this difference?

3.1. The well-known part

- This part of the hierarchy is relatively well-known in the literature:
- (11) questions > free relatives and correlatives > **“light-headed” relatives** > relatives with nouns in the head
- Lehmann (1984: 326): If a relative pronoun coincides with an interrogative pronoun, it always does so in (non-specific) free relatives (cf. also Caponigro 2002);
 - Corresponds to a well-established grammaticalization path (Lehmann 1984);
 - Lehmann (1984): free relatives often share the relative pronoun with (some) relative clauses with pronouns in the head;
 - The “intermediate” class with pronominal heads:
 - **“light-headed”** relatives (Citko 2004), the term is adopted below;
 - false free relatives (de Vries 2002).

3.2. The distinctions between the “light-headed” relatives

- ? Which properties of the head and the clause really matter?

3.2.1. Specificity and related ideas

- A difference more or less independently claimed in numerous works discussing “light-headed relatives” (Křížková 1970; Lehmann 1984; Spencer 1993):
 - relatives with non-specific reference
 - vs.
 - relatives with specific definite reference.
- Křížková 1970: in some Slavic languages, ‘which’ is more acceptable with the head ‘that’ if the relative construction has specific definite reference.

¹⁰ The differences between ‘that’ and ‘all’; ‘all’ and ‘those’ are statistically significant, χ^2 , $P < 0,05$.

¹¹ The differences between ‘that’ and ‘all’; ‘all’ and ‘those’ are statistically significant, Fisher’s exact test, $P < 0,05$.

- Specific reference could be more frequent in the plural. Could it account for the observed differences between numbers?
 - ‘that’, non-specific

(12) RUSSIAN, Denis Fonvizin, 1788 (RNC)

Basn’ učit, čto tot, kotoryj pervee vsej prinimaet mody,
 fable teaches that that which earlier than everyone accepts vogue

i tot, kotoryj deržitsja stariny, — oba ravnye duraki.

and that which keeps to the old both equal fools

‘The fable teaches us that the one who is the first to accept the new fashion and the one who keeps to the old are equally fools’.

- ‘that’, specific

(13) RUSSIAN, Nikolai Karamzin, 1793 (RNC)

Esli on sam poslal tebja — tot, kotorogo strašnoe prokljatie
 if he himself sent you that which.GEN terrifying curse

gremit vseгда v moem sluxe...

thunders always in my hearing

‘If he sent you himself, he whose terrible curse always sounds in my ears...’

- The results:

Table 7. Frequency of ‘who’ and ‘which’ in Russian of the 18th century, more details (RNC)

| | | ‘who’ | ‘which’ | ratio of ‘who’ |
|---------|--------------|-------|------------------|----------------|
| ‘that’ | non-specific | 79 | 21 | 0,79 |
| | specific | 29 | 71 ¹² | 0,29 |
| ‘all’ | | 3 | 7 | 0,30 |
| ‘those’ | | 6 | 161 | 0,04 |

- Statistically significant differences:
 - specific vs. non-specific relative constructions with ‘that’;
 - specific ‘that’ vs. ‘those’.
- In other words,
 - specificity does matter for this distinction in the Russian of the 18th century;
 - it can not explain on its own the difference between ‘that’ and ‘these’.

3.2.2. Semantic types of relative clauses

- The specificity distinction could probably be rephrased as that between maximalizing and restrictive relative clauses (for this distinction, see Grosu, Landman 1998);
- A further distinction is drawn in Russian, where ‘who’ is possible in restrictive relative clauses and impossible in non-restrictive relative clauses with nouns in the head;
- Ergo:

(14) appositive < restrictive < maximalizing

3.2.3. Number

- The middle of the hierarchy can probably be captured by two hierarchies:

(15) no (semantically regular) SG/PL distinction > SG/PL distinction

(16) SG > PL

¹² About half of the examples in this group have reference to the God. This subgroup has an even stronger tendency to contain the pronoun ‘which’.

- Browne W. (1986). *Relative Clauses in Serbo-Croatian in Comparison with English*. Zagreb.
- Caponigro I. (2002). Free relatives as DPs with silent D and a CP complement. In Vida Samiian (ed.). *Proceedings of WECOL 2000*. Fresno, Cal: California State University.
- Citko B. (2004). On headed, headless, and light-headed relatives. In *Natural Language & Linguistic Theory* 22, 1. P. 95–126.
- Faßke H. (1981). *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie*. Bautzen: Domowina.
- Gołąb Z., Friedman V. A. (1972). The relative clause in Slavic. In Peranteau P., Levi J., & Phares G. (eds.). *The Chicago Which Hunt. Papers from the Relative Clause Festival*. Chicago. P. 30–46.
- Grosu A., Landman F. (1998). Strange relatives of the third kind. In *Natural Language Semantics*, 6. P. 125–170.
- Itkonen T. (1997). *Kieliopas*. Helsinki: Kirjayhtymä. Non vidi.
- Kholodilova M. (2015). *To, čto nazyvaetsja: meždu odnositel'noj konstrukcij i ustojčivym vvodnym oborotom*. Talk at “Constructional & Lexical Semantic Approaches to Russian (IV)”. Saint Petersburg, 16.–18.04.15.
- Kholodilova M. (in print). Soglasovanie s veršinoj odnositel'nyx konstrukcij i obosoblennyx oborotov v russkom jazyke. In *Russkij jazyk v naučnom osveščeenii*.
- Kholodilova M. & Privizentseva M. (2015). Inverse attraction in Finno-Ugric languages. Talk at ‘Insufficient strength to defend its case’: Case attraction and related phenomena, Wrocław, 18–19 September 2015.
- Křížková H. (1970). Relativní věty v současných slovanských jazycích. In *Slávia* 39. P. 10–40.
- Lander Ju. A. (2011). Tipologija odnositel'nyx konstrukcij i pozicija semantičeskoj veršiny. Talk, Moscow, 2011.
- Lehmann, Ch. (1984). *Der Relativsatz: Typologie seiner Strukturen; Theorie seiner Funktionen; Kompendium seiner Grammatik*. Tübingen: Narr.
- Spencer D. (1993). *Aspects of the Syntax of Relative Clauses in Colloquial and Standard Russian*. PhD Thesis.
- de Vries M. (2002). *The Syntax of Relativization*. PhD Thesis.